

UNSUPERVISED CLASSIFICATION OF REMOTELY SENSED IMAGES USING GAUSSIAN MIXTURE MODELS AND PARTICLE SWARM OPTIMIZATION

Çağlar Arı

Selim Aksoy

Department of Electrical and Electronics Engineering
Bilkent University
Bilkent, 06800, Ankara, Turkey
cari@ee.bilkent.edu.tr

Department of Computer Engineering
Bilkent University
Bilkent, 06800, Ankara, Turkey
saksoy@cs.bilkent.edu.tr

ABSTRACT

Gaussian mixture models (GMM) are widely used for unsupervised classification applications in remote sensing. Expectation-Maximization (EM) is the standard algorithm employed to estimate the parameters of these models. However, such iterative optimization methods can easily get trapped into local maxima. Researchers use population-based stochastic search algorithms to obtain better estimates. We present a novel particle swarm optimization-based algorithm for maximum likelihood estimation of Gaussian mixture models. The proposed approach provides solutions for important problems in effective application of population-based algorithms to the clustering problem. We present a new parametrization for arbitrary covariance matrices that allows independent updating of individual parameters during the search process. We also describe an optimization formulation for identifying the correspondence relations between different parameter orderings of candidate solutions. Experiments on a hyperspectral image show better clustering results compared to the commonly used EM algorithm for estimating GMMs.

Index Terms— Gaussian mixture models, maximum likelihood estimation, particle swarm optimization, stochastic search, covariance parametrization, clustering

1. INTRODUCTION

Unsupervised classification, also called clustering, has been a classical problem in pattern recognition. Clustering has also been used for a wide variety of tasks in remote sensing image analysis such as pre-processing, segmentation, feature extraction, dimensionality reduction, data visualization, and final classification. One of the most widely used family of clustering algorithms includes iterative partitioning methods among which k-means and its extensions such as fuzzy c-means, k-medoids, and isodata have been the most popular choices.

These methods attempt to iteratively minimize an error criterion and terminate the iterations when a local minimum is reached.

Despite their popularity, these methods, and similarly most other clustering methods, have common problems in the following issues: restrictions on the shapes of the clusters, high dimensionality of the feature space, feature selection, identification of the number of clusters, and sensitivity to initialization. For example, the k-means algorithm that minimizes the sum of squared errors criterion is very limited in terms of its cluster modeling capability because it can only model spherical clusters with similar number of data points. It also gives equal importance to all features by using the Euclidean distance for point dissimilarity. Furthermore, the notion of distance in high dimensions becomes unclear when the feature space becomes very sparse compared to the number of available points. The model-based clustering approach that uses Gaussian mixture models (GMM) that are learned by maximizing the likelihood function using the expectation-maximization (EM) algorithm is superior to k-means in the sense that it is capable of finding clusters of arbitrary ellipsoidal shapes with arbitrary number of data points. However, significant difficulties in the estimation of the parameters of the GMM model (e.g., covariance matrix estimation) are observed in increasing dimensions. Furthermore, both the k-means and the GMM-EM algorithm are very sensitive to initializations and easily get trapped in local minima. In practice, these algorithms are run many times with different initial parameters, and various local search heuristics are used to find better parameters near the converged ones.

Constant increase in computational power has made population-based stochastic search algorithms very popular. Consequently, various population-based global optimization algorithms have been proposed to solve clustering problems. For example, Chang et al. [1] used a genetic algorithm to improve the k-means clustering algorithm, Maulik and Saha [2] used differential evolution for fuzzy c-means clustering, and Paoli et al. [3] used particle swarm optimization for estimating GMMs. All three methods were applied to pixel-based

This work was supported in part by the TUBITAK CAREER grant 104E074.

classification of satellite images.

The use of the Gaussian distribution as the class-conditional density model for multispectral data has been well accepted in the remote sensing literature. Therefore, it is of great interest to extend the population-based optimization algorithms for the estimation of GMMs. In this paper, we propose a clustering algorithm that uses particle swarm optimization (PSO) for finding an optimum solution for GMM estimation. The proposed algorithm solves three important problems that exist in related work: the lack of a suitable parametrization for arbitrary covariance matrices, updating of the parameters from data in conjunction with the stochastic search, and the degeneracy problem due to the interchangeability of different parameter orderings for the same candidate solution. Section 2 summarizes the general PSO framework and discusses the limitations of existing approaches, Section 3 presents the proposed clustering algorithm, and Section 4 illustrates its effectiveness in the classification of a hyperspectral image.

2. PARTICLE SWARM OPTIMIZATION

PSO is a population-based stochastic search algorithm based on the movement and the intelligence of swarm animals. In PSO, each solution is represented as a particle in a swarm. Each particle has a position vector z_i and velocity vector v_i . For a d -dimensional optimization problem, the position of each particle $z_i \in R^d$ represents a candidate solution. A fitness function uses the particle's position and assigns a fitness value to that particle. The particle having the best fitness value is called the global best (z_{gb}). Each particle also remembers its best position since the first iteration and this position is called the personal best ($z_{pb,i}$). In the first iteration, each particle is typically initialized with a random position and velocity. In the following iterations, each of the d velocity components in v_i is updated independently using the global best and its own personal best in a stochastic manner as

$$v_i(t+1) = w v_i(t) + c_1 r_1(t) (z_{pb,i}(t) - z_i(t)) + c_2 r_2(t) (z_{gb}(t) - z_i(t)) \quad (1)$$

where w is the inertia weight, r_1 and r_2 represent random numbers sampled from Uniform[0, 1], and c_1 and c_2 are small constants. The particle moves from its old position to a new position using its velocity vector as

$$z_i(t+1) = z_i(t) + v_i(t+1), \quad (2)$$

and updates its personal best if needed. After each iteration, the global best of the whole swarm is also updated.

The most important property of PSO is its use of the global best to coordinate the movement of all particles and the use of personal bests to remember the history of each particle where the global best serves as the current state of the problem and the personal bests serve as the current states of the particles. However, there are various problems that need

to be solved in order to effectively apply population-based algorithms like PSO to clustering problems. An important problem is the lack of a suitable parametrization for arbitrary covariance matrices. Since each component in the particle position vector z_i is independently updated using the corresponding component in the velocity vector v_i , it is not possible to include an arbitrary covariance matrix with $d(d+1)/2$ parameters to a particle definition because independent updates of the covariance components will very often violate the requirement for the matrix being symmetric and positive definite. Hence, existing methods limit their cluster model to diagonal covariance matrices [3] or do not use any covariance structure at all [1, 2]. We propose a new parametrization where the parameters of arbitrary covariance matrices are unique and independently modifiable.

Another problem is the updating of the parameters from data in addition to the randomized search procedure that can have convergence problems especially when the model contains many clusters with many parameters. The proposed covariance parametrization allows us to update the cluster parameters from data in conjunction with the stochastic search, and enables faster and more effective convergence.

The third problem that is tackled in this paper is the degeneracy problem. Degeneracy occurs when multiple representations for the same solution exist. There exists $K!$ different particle representations in clustering problems with K clusters due to different parameter orderings for the same candidate solution. This problem is often ignored in local search algorithms but it causes big problems for population-based stochastic search algorithms because the correspondences between cluster parameters of different particles are not known and particle updates using (1) may be based on wrong interactions. We propose a matching algorithm for finding the correct correspondences between the components of a particle and the global best for correct updates.

3. PROPOSED CLUSTERING ALGORITHM

The details of the proposed clustering algorithm that uses PSO for the estimation of GMMs are described below. The solution consists of a new parametrization where the components of full covariance matrices are unique and independently modifiable during stochastic search, a method for updating candidate solutions using data in conjunction with the search, and a formulation for ordering the parameters for finding correct correspondences between clusters during parameter updates. We also briefly describe the details of the initialization procedure.

Parametrization and particle definition: The proposed clustering model uses a mixture of K multivariate Gaussian distributions parametrized using $\Theta = \{\pi_1, \mu_1, \Sigma_1, \dots, \pi_K, \mu_K, \Sigma_K\}$ where $\pi_k \in [0, 1]$ are the prior probabilities, $\mu_k \in R^d$ are the means, and $\Sigma_k \in R^{d \times d}$ are the covariance matrices for the clusters $k = 1, \dots, K$. The K prior

probability values are calculated from the probabilistic assignments of the data points to the clusters, and are not part of the PSO particles. Each particle consists of the parameters of the mean vectors and the covariance matrices. Each mean vector $\mu \in R^d$ is parametrized by d real numbers. Our parametrization for each covariance matrix $\Sigma \in R^{d \times d}$ is based on its eigenvalue decomposition $\Sigma = V\Lambda V^T$ using the cyclic Jacobi algorithm and QR factorization of the corresponding eigenvector matrix $V = \prod_{1 \leq p < q \leq d} G(p, q, \phi^{pq})R$ via Givens rotation matrices $G(p, q, \phi^{pq})$ for $1 \leq p < q \leq d$ and a diagonal matrix R with entries being either +1 or -1. The eigenvalues are parametrized as d positive real numbers. Since the eigenvector matrix can be written as a multiplication of $d(d-1)/2$ Givens rotation matrices $G(p, q, \phi^{pq})$ where the angles ϕ^{pq} lie in the interval $[-\pi/2, +\pi/2]$ [4], it is parametrized in terms of the $d(d-1)/2$ Givens rotation angles. Robust estimation of covariance matrices in high dimensions becomes possible because regularization can be performed on very small eigenvalues to avoid singularities. This parametrization allows independent updating of each parameter and enables the use of full covariance models in the GMM.

Initialization: First, the K mean vectors are randomly selected from the data points. Then, the initial clusters are formed by assigning each data point to the closest mean. Finally, the covariance matrix of each cluster is computed and the angles and eigenvalues are estimated using the cyclic Jacobi algorithm and QR factorization [4].

Optimization: The PSO iterations proceed to find the parameters that minimize the negative log-likelihood function with the assumption that the data points are independent and identically distributed.

Parameter ordering: Before updating each particle using (1) and (2), the correspondences between its clusters and the clusters of the global best particle are found. The matching problem is formulated as a minimum cost network flow optimization problem where the objective is to find an ordering of individual clusters so that the sum of Mahalanobis distances between the means of a particle and the corresponding means of the global best particle is minimized.

Here $\{\mu_{X_{PB}}^{(i)}\}_{i=1}^K$ and $\{\Sigma_{X_{PB}}^{(i)}\}_{i=1}^K$ represent the set of personal best means and covariance matrices of a particle of interest, and similarly $\{\mu_{X_{GB}}^{(j)}\}_{j=1}^K$ and $\{\Sigma_{X_{GB}}^{(j)}\}_{j=1}^K$ represent the set of means and covariance matrices for the global best particle. The cost of matching the former particle's i 'th cluster parameters to the global best particle's j 'th cluster parameters is computed as

$$c_{ij} = (\mu_{X_{PB}}^{(i)} - \mu_{X_{GB}}^{(j)})^T (\Sigma_{X_{GB}}^{(j)})^{-1} (\mu_{X_{PB}}^{(i)} - \mu_{X_{GB}}^{(j)}), \quad (3)$$

and the correspondences are found using the Edmonds-Karp

algorithm [5] that solves the following optimization problem:

$$\begin{aligned} & \underset{I_{11}, \dots, I_{KK}}{\text{minimize}} && \sum_{i=1}^K \sum_{j=1}^K c_{ij} I_{ij} \\ & \text{subject to} && \sum_{i=1}^K I_{ij} = 1, \quad \forall j \in \{1, \dots, K\} \\ & && \sum_{j=1}^K I_{ij} = 1, \quad \forall i \in \{1, \dots, K\} \\ & && I_{ij} = \begin{cases} 1, & \text{correspondence between} \\ & i\text{'th and } j\text{'th clusters} \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \quad (4)$$

Update equations: The correspondence relation computed above is denoted with a function $f(k)$ that maps the current particle's cluster index k to the corresponding global best particle's cluster index $f(k)$. Mean and covariance parameters of particles are updated by using correct correspondence relations as follows:

Mean update equations

$$\mu_{V_{t+1}}^{(k)} = w\mu_{V_t}^{(k)} + c_1(\mu_{PB_t}^{(k)} - \mu_{X_t}^{(k)}) + c_2(\mu_{GB_t}^{(f(k))} - \mu_{X_t}^{(k)})$$

$$\mu_{X_{t+1}}^{(k)} = \mu_{X_t}^{(k)} + \mu_{V_{t+1}}^{(k)}$$

Covariance update equations — Angle updates

$$\phi_{V_{t+1}}^{pq,(k)} = w\phi_{V_t}^{pq,(k)} + c_1(\phi_{PB_t}^{pq,(k)} - \phi_{X_t}^{pq,(k)}) + c_2(\phi_{GB_t}^{pq,(f(k))} - \phi_{X_t}^{pq,(k)})$$

$$\phi_{X_{t+1}}^{pq,(k)} = \phi_{X_t}^{pq,(k)} + \phi_{V_{t+1}}^{pq,(k)}$$

Covariance update equations — Eigenvalue updates

$$\lambda_{V_{t+1}}^{i,(k)} = w\lambda_{V_t}^{i,(k)} + c_1(\lambda_{PB_t}^{i,(k)} - \lambda_{X_t}^{i,(k)}) + c_2(\lambda_{GB_t}^{i,(f(k))} - \lambda_{X_t}^{i,(k)})$$

$$\lambda_{X_{t+1}}^{i,(k)} = \lambda_{X_t}^{i,(k)} + \lambda_{V_{t+1}}^{i,(k)}$$

Particle updating using data:

The parameters of the global best particle are updated by estimating a new covariance matrix using the data points assigned to each cluster and by performing another set of eigenvalue decomposition $\Sigma = V\Lambda V^T$ and QR factorization $V = \prod_{1 \leq p < q \leq d} G(p, q, \phi^{pq})R$ steps.

4. EXPERIMENTS

The experiments were performed to compare the proposed PSO-based clustering algorithm with the GMM-EM algorithm using a 145×145 pixel AVIRIS image taken over Indiana's Indian Pines test site. Since the GMM-EM algorithm cannot estimate full covariance matrices due to singularity issues, the 9-band subset that came with the original data was used instead of the whole set of 220 bands.

Figures 1(a) and 1(b) show the false color image and the 16-class ground truth, respectively. As the best possible performance that can be achieved by Gaussian classification to serve as the baseline for comparison, we performed supervised maximum likelihood classification using the whole ground truth as training data as shown in Figure 1(c).

Both PSO and GMM-EM were initialized by randomly selecting K mean vectors from the data points and the initial clusters were generated by assigning each data point to the closest mean. In each experiment, GMM-EM and PSO algorithms both used the same initializations. At the end of each

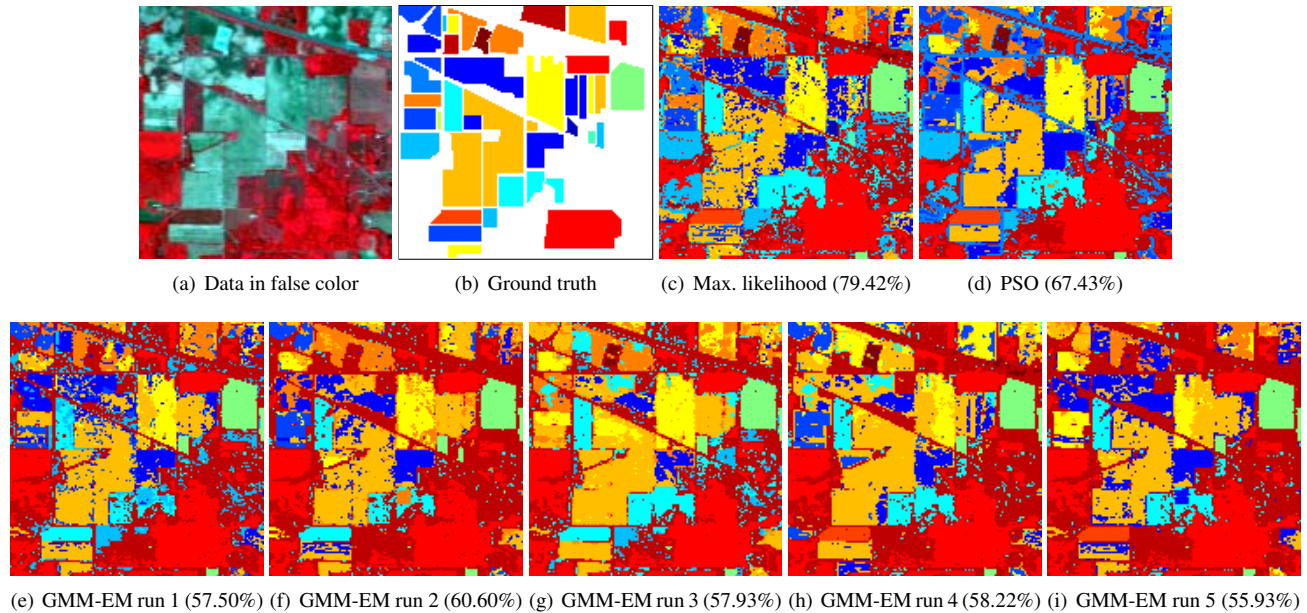


Fig. 1. Results for the Indian Pines data set. Classification accuracies are given in parenthesis.

experiment, the parameters corresponding to the global best particle were used as the result of the PSO algorithm, and the parameters of the GMM-EM result having the highest likelihood value among different initializations were used as the GMM-EM result.

Figure 1 shows the results of a particular run of the PSO algorithm and five GMM-EM runs for comparison. The PSO algorithm was run using 30 particles for 60 iterations with the results shown in Figure 1(d). The GMM-EM procedure was run using 30 different random initializations for 500 iterations. Figures 1(e)–1(i) show the results for five cases that resulted in the highest likelihood values. Quantitative results computed by matching the ground truth labels with the clusters obtained using the unsupervised methods showed that the proposed algorithm obtained an accuracy rate closer to the one by the supervised classifier compared to the best GMM-EM runs.

5. CONCLUSIONS

The use of the Gaussian distribution as the density model for multispectral data in both unsupervised and supervised classification problems has been well accepted in the remote sensing literature. We presented a stochastic search-based approach that used particle swarm optimization for finding the maximum likelihood estimates of Gaussian mixture models for clustering multivariate data. The proposed algorithm provided solutions for three important problems in effective application of population-based algorithms like PSO to this clustering problem: the lack of a suitable parametrization for arbitrary covariance matrices, updating of the parameters

from data in conjunction with the stochastic search, and the degeneracy problem due to the interchangeability of different parameter orderings for the same candidate solution. Experiments on a hyperspectral image showed better clustering results compared to the commonly used EM algorithm for estimating GMMs. Possibilities for future work include adding heuristics for faster convergence of the PSO iterations and adding constraints for feature selection.

6. REFERENCES

- [1] D.-X. Chang, X.-D. Zhang, and C.-W. Zheng, “A genetic algorithm with gene rearrangement for k-means clustering,” *Pattern Recognition*, vol. 42, no. 7, pp. 1210–1222, July 2009.
- [2] U. Maulik and I. Saha, “Modified differential evolution based fuzzy clustering for pixel classification in remote sensing imagery,” *Pattern Recognition*, vol. 42, no. 9, pp. 2135–2149, September 2009.
- [3] A. Paoli, F. Melgani, and E. Pasolli, “Clustering of hyperspectral images based on multiobjective particle swarm optimization,” *IEEE TGARS*, vol. 47, no. 12, pp. 4175–4188, December 2009.
- [4] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 3rd edition, 1996.
- [5] J. Edmonds and R. M. Karp, “Theoretical improvements in algorithmic efficiency for network flow problems,” *Journal of the ACM*, vol. 19, no. 2, pp. 248–264, April 1972.